

This Page Is Inserted by IFW Operations  
and is not a part of the Official Record

## **BEST AVAILABLE IMAGES**

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

**IMAGES ARE BEST AVAILABLE COPY.**

**As rescanning documents *will not* correct images,  
please do not report the images to the  
Image Problem Mailbox.**

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 18, 2003, 03:18:06 ; Search time 9.07541 Seconds  
(without alignments)  
1581.285 Million cell updates/sec

Title: US-09-807-933B-11  
Perfect score: 1895  
Sequence: 1 MKFSIIASALLAASSTYAA.....TFKAVTCPAEIIAKTCERK 346

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_40.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Query %	Length	ID	Description
1	651	34.4	213	1	GUN5 HUMIN	P43316 humicola in
2	644.5	34.0	376	1	GUNK FUSOX	P45699 fusarium ox
3	528.5	27.9	511	1	GUNB PSEFL	P18126 pseudomonas
4	461.5	24.4	393	1	GUN1 USTMA	P54424 ustilago ma
5	220.5	11.6	662	1	MUC1 XENLA	Q05049 xenopus lae
6	200.5	10.6	471	1	GUX3 TRIRE	P07987 trichoderma
7	192.5	10.2	438	1	GUX3 AGABI	P49075 agaricus bi
8	187	9.9	307	1	SGS3 DROME	P02840 drosophila
9	181.5	9.6	217	1	SGS3 DROSI	P13729 drosophila
10	181.5	9.6	263	1	SGS3 DROVA	P13728 drosophila
11	180	9.5	797	1	VGLX HSVEB	P28958 equine herp
12	177.5	9.4	418	1	GUN3 TRIRE	P07982 trichoderma
13	164	8.7	5179	1	MUC2 HUMAN	Q02817 homo sapien
14	163	8.6	1161	1	DANA YEAST	P47179 saccharomyc
15	162.5	8.6	400	1	MUAD XENLA	P10667 xenopus lae
16	162.5	8.6	802	1	XYNL RUMFL	Q53317 ruminococcu
17	162	8.5	3178	1	YS89 CAEEL	Q09624 caenorhabdi
18	156.5	8.3	388	1	GUN5 HUMIN	Q12624 humicola in
19	156	8.2	462	1	GUNB FUSOX	P46236 fusarium ox
20	153	8.1	183	1	AAC1 DICDI	P14195 dictyosteli
21	150	7.9	275	1	UL11 HCMVA	P16721 human cytom
22	149	7.9	600	1	SP96 DICDI	P14328 dictyosteli
23	148	7.8	617	1	CHIT CAEEL	Q11174 caenorhabdi
24	147	7.8	725	1	AGAI YEAST	P32323 saccharomyc
25	146.5	7.7	304	1	YQOB CAEEL	Q09300 caenorhabdi
26	145.5	7.7	587	1	HE PARLI	P22757 paracentrot
27	144.5	7.6	1322	1	YAG3 YEAST	P39712 saccharomyc
28	144	7.6	624	1	YH19 YEAST	P38900 saccharomyc
29	143.5	7.6	605	1	YHC8 YEAST	P38739 saccharomyc
30	143.5	7.6	786	1	STUB DROME	Q05319 drosophila
31	140.5	7.4	329	1	YK23 YEAST	P36110 saccharomyc
32	140	7.4	1374	1	YMN3 YEAST	Q03099 saccharomyc
33	140	7.4	1537	1	FLOI YEAST	P32768 saccharomyc

ALIGNMENTS

RESULT 1				
ID	GUN5_HUMIN	STANDARD;	PRT;	213 AA.
AC	P43316;			
DT	01-NOV-1995 (Rel. 32, Created)			
DT	01-NOV-1995 (Rel. 32, Last sequence update)			
DT	15-DEC-1998 (Rel. 37, Last annotation update)			
DE	Endoglucanase V (EC 3.2.1.4) (Endo-1,4-beta-glucanase V)			
DE	(Cellulase V) (EG V).			
OS	Humicola insolens.			
OC	Eukaryota; Fungi; Ascomycota; mitosporic Ascomycota; Humicola.			
OX	NCBI_TaxID=34413;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RA	Rasmussen G., Mikkelsen J.-M., Schulein M., Patkar S.A., Hagen F.,			
RA	Hjort C.M., Hastrup S.;			
RT	"A cellulase preparation comprising an endoglucanase enzyme.";			
RL	Patent number WO9117243, 14-NOV-1991.			
RN	[2]			
RP	X-RAY CRYSTALLOGRAPHY (1.6 ANGSTROMS).			
RX	MEDLINE=93390621; PubMed=8377830;			
RA	Davies G.J., Dodson G.G., Hubbard R.E., Tolley S.P., Dauter Z.,			
RA	Wilson K.S., Hjort C., Mikkelsen J.M., Rasmussen G., Schulein M.;			
RT	"Structure and function of endoglucanase V.";			
RL	Nature 365:362-364(1993).			
RN	[3]			
RP	X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS).			
RX	MEDLINE=96101453; PubMed=8519779;			
RA	Davies G.J., Tolley S.P., Henriessat B., Hjort C., Schulein M.;			
RT	"Structures of oligosaccharide-bound forms of the endoglucanase V			
RL	from Humicola insolens at 1.9-A resolution.";			
RN	Biochemistry 34:16210-16220(1995).			
RN	[4]			
RP	X-RAY CRYSTALLOGRAPHY (1.5 ANGSTROMS).			
RA	Davies G.J., Dodson G.G., Moore M.H., Tolley S.P., Dauter Z.,			
RA	Wilson K.S., Rasmussen G., Schulein M.;			
RT	"Structure determination and refinement of the Humicola insolens			
RL	endoglucanase V at 1.5-A resolution.";			
RN	Acta Crystallogr. D 52:7-17(1996).			
CC	-1- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic			
CC	linkages in cellulose.			
CC	-1- SIMILARITY: BELONGS TO CELLULASE FAMILY K (FAMILY 45 OF GLYCOSYL			
CC	HYDROLASES).			
DR	PDB; 2ENG; 08-DEC-96.			
DR	PDB; 3ENG; 16-JUN-97.			
DR	PDB; 4ENG; 16-JUN-97.			
DR	InterPro; IPR000334; GH 45.			
DR	Pfam; PF02015; Glyco_hydro_45; 1.			
DR	PROSITE; PS01140; GLYCOSYL_HYDROL_F45; 1.			
KW	Cellulose degradation; Hydrolase; Glycosidase; 3D-structure.			
FT	ACT_SITE 10 10 NUCLEOPHILE.			
FT	ACT_SITE 121 121 PROTON DONOR.			
SQ	SEQUENCE 213 AA; 22864 MW; 24334301BA3BC804 CRC64;			
Query Match 34.4%; Score 651; DB 1; Length 213;				
Best Local Similarity 57.1%; Pred. No. 1.7e-39;				

P40105 saccharomyc  
P53345 saccharomyc  
P53819 saccharomyc  
P13730 drosophila  
O13559 saccharomyc  
P40434 saccharomyc  
P40889 saccharomyc  
P24088 saccharomyc  
P38894 saccharomyc  
P04065 saccharomyc  
P16833 human cytom  
Q01836 listeria in

Matches 116; Conservative 24; Mismatches 59; Indels 4; Gaps 2;	
QY 144	NGRTRTYWDCCKPSCAWDGHKASVTKPVLTKAKDGVSRGLSDVOSGC-VGGQAYMCDNQP 202
DB 2	DGRSTRYWDCKPSCGWAAPVNPQPVFSCNANFQRTIDPDAKSGCEPGGVAYSCADQTP 61
QY 203	WVNDLLAYGFAAASLGASAGAFCCGCGCYELTFTNTAVAGKFFVQVNTGDDLSSTNHFD 262
DB 62	WVNDLLAYGFAAASLGASAGAFCCGCGCYELTFTNTAVAGKFFVQVNTGDDLSSTNHFD 121
QY 263	LQMPGGGVGVYFNGCQSQWNTNTDQWARYGGISSECDKLPQLQAGCKWRFGWFKNAD 322
DB 122	LNIPGGGVGVYFNGCQSQWNTNTDQWARYGGISSECDKLPQLQAGCKWRFGWFKNAD 178
QY 323	NPEVTFKAVTCPAEIIAKTGCR 345
DB 179	NPSFQRQVQCPAELVARTGCR 201
RESULT 2	
ID	GUNK_FUSOX
AC	P45699;
DT	01-NOV-1995 (Rel. 32, Created)
DT	01-NOV-1995 (Rel. 32, Last sequence update)
DT	01-NOV-1995 (Rel. 32, Last annotation update)
DE	Putative endoglucanase type K precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase) (Cellulase).
DE	Fusarium oxysporum.
OS	Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC	Hypocreales; mitosporic Hypocreales; Fusarium.
OX	NCBI_TaxID=5507;
RN	[1]
RP	SEQUENCE FROM N.A.
RA	MEDLINE=95047531; PubMed=7959045;
RA	Sheppard P.O., Grant F.J., Oort F.J., Sprecher C.A., Foster D.C.,
RA	Hagen F.S., Upshall A., McKnight G.L., O'Hara P.J.;
RT	"The use of conserved cellulase family-specific sequences to clone
RT	cellulase homologue cDNAs from Fusarium oxysporum.;"
RL	Gene 150:163-167(1994).
CC	-!- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic
CC	linkages in cellulose.
CC	-!- SIMILARITY: CONTAINS 1 FUNGAL-TYPE CELLULOSE-BINDING DOMAIN (CBD).
CC	-!- SIMILARITY: BELONGS TO CELLULOSE FAMILY K (FAMILY 45 OF GLYCOSYL
CC	HYDROLASES).
This SWISS-PROT entry is copyright. It is produced through a collaboration	
between the Swiss Institute of Bioinformatics and the EMBL outstation -	
the European Bioinformatics Institute. There are no restrictions on its	
use by non-profit institutions as long as its content is in no way	
modified and this statement is not removed. Usage by and for commercial	
entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>	
or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).	
-----	
DR	EMBL; L29381; AAA65589.1; -
DR	HSSP; P43316; 2ENG.
DR	InterPro; IPR000254; CBD fungal.
DR	InterPro; IPR000334; GH_45.
DR	Pfam; PF00734; CBM_1; 1.
DR	Pfam; PF02015; Glyco_hydro_45; 1.
DR	SMART; SM00236; fCBD; 1.
DR	PROSITE; PS00562; CBD_FUNGAL; 1.
DR	PROSITE; PS01140; GLYCOSYL_HYDROL_F45; 1.
KW	Cellulose degradation; Hydrolase; Glycosidase; Signal.
FT	SIGNAL 1 18 POTENTIAL.
FT	CHAIN 19 376 PUTATIVE ENDOGLUCANASE TYPE K.
FT	DOMAIN 19 308 CATALYTIC.
FT	DOMAIN 309 338 LINKER.
FT	DOMAIN 339 376 CELLULOSE-BINDING.
FT	ACT SITE 29 29 NUCLEOPHILE (BY SIMILARITY).
FT	ACT SITE 140 140 PROTON DONOR (BY SIMILARITY).
SQ	SEQUENCE 376 AA; 39235 MW; B430ASF962B9F882 CRC64;

Query Match 34.0%; Score 644.5; DB 1; Length 376;	
Best Local Similarity 55.1%; Pred. No. 8.3e-39;	
Matches 113; Conservative 27; Mismatches 62; Indels 3; Gaps 2;	
QY 142	SGNRTRTYWDCCKPSCAWDGHKASVTKPVLTKAKDGVSRGLSDVOSGCV-GGQAYMCDNQ 200
DB 19	SGSGHSTRYWDCKPSCSWGSKAAVNAFALTCDKNDNPFISNTNNAVNGCEGGGSAYACTNY 78
QY 201	QPVNDLLAYGFAAASLGASAGAFCCGCGCYELTFTNTAVAGKFFVQVNTGDDLSSTNH 260
DB 79	SPWAVNDELAAYGFAATKISGGSEASWCCACAYALTFTTGPVKKKMIVQSTNTGGDLGDH 138
QY 261	FDLQMPGGGVGVYFNGCQSQWNTNTDQWARYGGISSECDKLPQLQAGCKWRFGWFKN 320
DB 139	FDLQMPGGGVGVYFNGCQSQWNTNTDQWARYGGISSECDKLPQLQAGCKWRFGWFKN 196
QY 321	ADNPEVTFKAVTCPAEIIAKTGCR 345
DB 197	ADNPDTTFEQVQCPKALLDISGCR 221
RESULT 3	
ID	GUNK_PSEFL
AC	P18126;
DT	01-NOV-1990 (Rel. 16, Created)
DT	01-NOV-1990 (Rel. 16, Last sequence update)
DT	15-DEC-1998 (Rel. 37, Last annotation update)
DE	Endoglucanase B precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase) (Cellulase) (EGB).
DE	CEL.
GN	CEL.
OS	Pseudomonas fluorescens.
OC	Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC	Pseudomonas
OX	NCBI_TaxID=294;
RN	[1]
RP	SEQUENCE FROM N.A., AND SEQUENCE OF 30-48.
RA	STRAIN=Sp. Cellulosa;
RA	MEDLINE=90355836; PubMed=2117693;
RA	Gilbert H.J., Hall J., Hazlewood G.P., Ferreira L.M.A.;
RT	"The N-terminal region of an endoglucanase from Pseudomonas
RT	fluorescens subspecies cellulosa constitutes a cellulose-binding
RT	domain that is distinct from the catalytic centre.;"
RL	Mol. Microbiol. 4:759-767(1990).
CC	-!- FUNCTION: THIS ENZYME CATALYZES THE ENDOHYDROLYSIS OF 1,4-BETA-
CC	GLUCOSIDIC LINKAGES IN CELLULOSE, LICHENIN AND CEREAL BETA-D-
CC	GLUCANS. EGB IS MOST ACTIVE AGAINST BARLEY BETA-GLUCAN, BUT SHOWED
CC	SIGNIFICANT ACTIVITY AGAINST AMORPHOUS AND CRISTALLINE CELLULOSE.
CC	-!- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic
CC	linkages in cellulose.
CC	-!- SUBCELLULAR LOCATION: Periplasmic.
CC	-!- SIMILARITY: CONTAINS 1 BACTERIAL-TYPE CELLULOSE-BINDING DOMAIN
CC	(CBD).
CC	-!- SIMILARITY: BELONGS TO CELLULOSE FAMILY K (FAMILY 45 OF GLYCOSYL
CC	HYDROLASES).
This SWISS-PROT entry is copyright. It is produced through a collaboration	
between the Swiss Institute of Bioinformatics and the EMBL outstation -	
the European Bioinformatics Institute. There are no restrictions on its	
use by non-profit institutions as long as its content is in no way	
modified and this statement is not removed. Usage by and for commercial	
entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>	
or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).	
-----	
DR	EMBL; X52615; CAA36844.1; -
DR	PIR; S10527; S10527.
DR	HSSP; P43316; 2ENG.
DR	InterPro; IPR001919; Bac_celose-bind.
DR	InterPro; IPR002883; CBD_5.
DR	InterPro; IPR000334; GH_45.
DR	Pfam; PF00553; CBM_2; 1.
DR	Pfam; PF02013; CBM_10; 1.
DR	Pfam; PF02015; Glyco_hydro_45; 1.

```
DR PROSITE; PS00561; CBD BACTERIAL; 1.
DR PROSITE; PS01140; GLYCOSYL HYDROL_F45; 1.
KW Cellulose degradation; Hydrolase; Glycosidase; Signal; Periplasmic.
FT SIGNAL 1 29
FT CHAIN 30 511 ENDOGLUCANASE B.
FT DOMAIN 30 131 CELLULOSE-BINDING (BY SIMILARITY).
FT DOMAIN 132 173 SER-RICH (LINKER).
FT DOMAIN 223 259 SER-RICH.
FT DISULFID 32 127 BY SIMILARITY.
FT ACT_SITE 276 276 NUCLEOPHILE (BY SIMILARITY).
FT ACT_SITE 393 393 PROTON DONOR (BY SIMILARITY).
SQ SEQUENCE 511 AA; 52078 MW; 3C119D998291D8E CRC64;

Query Match 27.9%; Score 528.5; DB 1; Length 511;
Best Local Similarity 31.8%; Pred. No. 1.7e-30;
Matches 122; Conservative 65; Mismatches 136; Indels 61; Gaps 12;

QY 4 SIYASALLAASYAAECQYQCGGKMWGTCTCTGTCVGAENNEWQCIP--N 61
DB 136 SSVASS--SSSSSVSSSTPRSSSSSVSSVPCTSSSSSSVLTGAACNWTGLTPLCN 192
QY 62 DQVQG-----NPKTTTTTTTAAATTKAPVTTTATTTTAKPVTTTKATTTTITTKTK 117
DB 193 NTSNGWGYEDRGSCVARTTCAQAPAPYGVISTSSST-----PLSSSSSSRSVASSSSL 246
QY 118 TTTTAAATTTSSNTGVSPTSGSGFSGNGRTRTRVWDCCKPSCAWDGKA-SVTKPVLTKAKD 176
DB 247 SSAT--SSASVSVVPPIDGCG--NGVATRWDCCKPHCGMSNVPSLVSPQCSAN 301
QY 177 GVSRLGSDVQSGCGVGOAYMNCNDQPNVNDLALAYGFAAASLGSAGASAFCCCYELTFT 236
DB 302 NTRLSDVSVGSSCDGGGYMCDKIPFVSPPTLAYGVATSSGDV-----CGRCYQLQFT 356
QY 237 -----NTAVAGKKEFVVQVNTGDDLSTNHFDLQMPGGGVYFNGCQSQMNTND 285
DB 357 GSSYNAPGDPGSAALAGKTIVQATNIGYDVSGQFDILVPGGVGVAFNACSAQWGVNSA 416
QY 286 GWCARYGGI-----SSISE-----CDKL-----PTLOAGCKKRFQPKA 321
DB 417 ELGAQYGGFLAACKQQLGYNASVQYSKSVLNRCDVSFGSRGLTQLQQGCTWFAENFEAA 476
QY 322 DNPVTFKAVTCAPIIAKTGCR 345
DB 477 DNPFLKYKEVPCPAELTTRSGMNR 500

RESULT 4
GUNI_GUN1 USTMA STANDARD; PRT; 393 AA.
AC P54424;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DE 01-NOV-1997 (Rel. 35, Last annotation update)
DE Endoglucanase 1 precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase 1)
DE (Cellulase 1) (EG 1).
GN EGL1.
OS Ustilago maydis (Smut fungus).
OC Eukaryota; Fungi; Basidiomycota; Ustilaginomycetes;
OC Ustilaginomycetidae; Ustilaginales; Ustilaginaceae; Ustilago.
OX NCBI_TaxID=5270;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=FBD11;
RX MEDLINE=96145728; PubMed=8590631;
RA Schauwecker F., Wanner G., Kahmann R.;
RT "Filament-specific expression of a cellulase gene in the dimorphic
RT fungus Ustilago maydis."
RL Biol. Chem. Hoppe-Seyler 376:617-625(1995).
CC -1- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic
CC linkages in cellulose.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- TISSUE SPECIFICITY: HYPHAL TIP.
CC -1- DEVELOPMENTAL STAGE: EXPRESSED IN FILAMENTOUS DIKARYON.
```



RA Faergestad L.G., Pettersson L.G.;  
 RT "The 1,4-beta-glucan cellobiohydrolases of *Trichoderma reesei* QM  
 RT 9414.";  
 RN FEBS Lett. 119:97-100(1980).  
 RN [4]  
 RP X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS).  
 RX MEDLINE=9033255; PubMed=2377893;  
 RA Rouvinen J., Bergfors T., Teeri T.T., Knowles J.K.C., Jones T.A.;  
 RT "Three-dimensional structure of cellobiohydrolase II from *Trichoderma*  
 RT *reesei*.";  
 RL Science 249:380-386(1990).  
 RN [5]  
 RP X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS).  
 RX MEDLINE=97029636; PubMed=8875646;  
 RA Koivula A., Reinikainen T., Ruohonen L., Valkeajärvi A., Rouvinen J.,  
 RA Claeysmans M., Telman O., Kleywegt G.J., Szardenings M., Rouvinen J.,  
 RA Jones T.A., Teeri T.T.;  
 RT "The active site of *Trichoderma reesei* cellobiohydrolase II: the role  
 RT of tyrosine 169.";  
 RL Protein Eng. 9:691-699(1996).  
 CC -1- FUNCTION: THE BIOLOGICAL CONVERSION OF CELLULOSE TO GLUCOSE  
 CC GENERALLY REQUIRES THREE TYPES OF HYDROLYTIC ENZYMES:  
 CC (1) ENDGLUCANASES WHICH CUT INTERNAL BETA-1,4-GLUCOSIDIC BONDS;  
 CC (2) EXOCELLULOBIODOLASES THAT CUT THE DISSACCHARIDE CELLOBIOSE  
 CC FROM THE NONREDUCING END OF THE CELLULOSE POLYMER CHAIN;  
 CC (3) BETA-1,4-GLUCOSIDASES WHICH HYDROLYZE THE CELLOBIOSE AND OTHER  
 CC SHORT CELLO-OLIGOSACCHARIDES TO GLUCOSE.  
 CC -1- CATALYTIC ACTIVITY: Hydrolysis of 1,4-beta-D-glucosidic linkages  
 CC in cellulose and cellotetraose, releasing cellobiose from the non-  
 CC reducing ends of the chains.  
 CC -1- SUBCELLULAR LOCATION: Secreted.  
 CC -1- MISCELLANEOUS: T. REESEI PRODUCES TWO DIFFERENT  
 CC EXOCELLULOBIODOLASES. THEY ARE UNIQUE IN THAT THEY CAN HYDROLYSE  
 CC CRYSTALLINE CELLULOSE IN THE ABSENCE OF ENDOGLUCANASES.  
 CC -1- SIMILARITY: BELONGS TO CELLULOSE FAMILY B (FAMILY 6 OF GLYCOSYL  
 CC HYDROLASES).  
 CC -1- SIMILARITY: CONTAINS 1 FUNGAL-TYPE CELLULOSE-BINDING DOMAIN (CBD).  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC -----  
 DR EMBL; M16190; AAA34210.1; -;  
 DR EMBL; M55080; AAA72922.1; -;  
 DR PIR; A26472; A26472.  
 DR PIR; A26160; A26160.  
 DR PDB; 3CBH; 15-JAN-91.  
 DR PDB; 1CB2; 19-MAR-99.  
 DR InterPro; IPR000254; CBD fungal.  
 DR InterPro; IPR001524; GH\_6.  
 DR Pfam; PF00734; CBM 1; 1.  
 DR Pfam; PF01341; Glyco hydro 6; 1.  
 DR PRINTS; PR00733; GLHYDRLASE6.  
 DR ProDom; PD001821; CBD fungal; 1.  
 DR ProDom; PD003733; GH\_6; 1.  
 DR SMART; SM00236; fCB2; 1.  
 DR PROSITE; PS00562; CBD FUNGAL; 1.  
 DR PROSITE; PS00655; GLYCOSYL HYDROL\_F6\_1; 1.  
 DR PROSITE; PS00656; GLYCOSYL HYDROL\_F6\_2; 1.  
 KW Cellulose degradation; Hydrolase; Glycosidase; Glycoprotein; Signal;  
 FT 3D-structure. 1 24  
 FT SIGNAL 25 471 EXOGLUCANASE II.  
 FT CHAIN 25 471 CELLULOSE-BINDING (BY SIMILARITY).  
 FT DOMAIN 25 65 LINKER.  
 FT DOMAIN 66 106 CATALYTIC.  
 FT DOMAIN 107 471  
 FT ACT\_SITE 199 199  
 FT ACT\_SITE 245 245 PROTON DONOR.  
 FT ACT\_SITE 425 425 NUCLEOPHILE.

FT DISULFID 34 51 BY SIMILARITY.  
 FT DISULFID 45 61 BY SIMILARITY.  
 FT DISULFID 200 259  
 FT DISULFID 392 439  
 FT CARBOHYD 38 38 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 111 111 O-LINKED (MAN. . .).  
 FT CARBOHYD 121 121 O-LINKED (MAN. . .).  
 FT CARBOHYD 130 130 O-LINKED (MAN. . .).  
 FT CARBOHYD 133 133 O-LINKED (MAN. . .).  
 FT CARBOHYD 134 134 O-LINKED (MAN. . .).  
 FT CARBOHYD 139 139 O-LINKED (MAN. . .).  
 FT CARBOHYD 313 313 N-LINKED (GLCNAC. . .).  
 FT CARBOHYD 334 334 N-LINKED (GLCNAC. . .).  
 FT MUTAGEN 199 199 D->A: 20% OF WILD-TYPE ACTIVITY.  
 FT MUTAGEN 245 245 D->A: NO MEASURABLE ACTIVITY.  
 FT CONFLICT 359 359 P -> R (IN REF. 2).  
 FT CONFLICT 449 449 P -> A (IN REF. 2).  
 SQ SEQUENCE 471 AA; 49653 MW; C4711BC35B1B08 CRC64;  
 Query Match 10.6%; Score 200.5; DB 1; Length 471;  
 Best Local Similarity 32.4%; Pred. No. 2.4e-07;  
 Matches 48; Conservative 26; Mismatches 41; Indels 33; Gaps 6;  
 QY 5 ITASALLLNASSTYAAE-----CSQYGCQGGKMTGPTCTCTSGFTCVGAENNEWYS 56  
 Db 2 IVGILTTIATLTAASVLEERQACSSVWGCGQWWSGPTCCASGTCV--YSNDYYS 59  
 QY 57 QCIPNDQVGNPKTTTTTKAATTTKAPVTTTKATTTTTPVTTTKATTTT-KTT 115  
 Db 60 QCLFG-----AASS-----SSTRAASSTTSRVSPPTSSSSATPPPGSGTT 99  
 QY 116 TKTTTTKAATTTSSSN--TCYSPISGGF 141  
 Db 100 TRVPPVSGTATYSGNPFVGVTPWANAY 127  
 RESULT 7  
 ID\_GUX3 AGABI STANDARD; PRT; 438 AA.  
 AC P49075;  
 DT 01-FEB-1996 (Rel. 33, Created)  
 DT 01-FEB-1996 (Rel. 33, Last sequence update)  
 DT 15-JUL-1999 (Rel. 38, Last annotation update)  
 DE Exoglucanase 3 precursor (EC 3.2.1.91) (Exocellobiohydrolase 3)  
 DE (1,4-beta-cellobiohydrolase 3).  
 GN CEL3.  
 OS Agaricus bisporus (Common mushroom).  
 OC Eukaryota; Fungi; Basidiomycota; Hymenomycetes; Homobasidiomycetes;  
 OC Agaricales; Agaricaceae; Agaricus.  
 OX NCBI\_TaxID=5341;  
 RN [1]  
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 255-277 AND 331-351.  
 RX STRAIN=D649;  
 RX MEDLINE=94368092; PubMed=8085821;  
 RA Chow C.-M., Yague E., Raguz S., Wood D.A., Thurston C.F.;  
 RT "The cel3 gene of *Agaricus bisporus* codes for a modular cellulase and  
 RT is transcriptionally regulated by the carbon source.";  
 RL Appl. Environ. Microbiol. 60:2779-2785(1994).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RP TISSUE=Mycelium;  
 RX MEDLINE=96269930; PubMed=8662210;  
 RA Yague E., Chow C.-M., Challen M.P., Thurston C.F.;  
 RT "Correlation of exons with functional domains and folding regions in  
 RT a cellulase from *Agaricus bisporus*.";  
 RL Curr. Genet. 30:56-61(1996).  
 CC -1- FUNCTION: SHOWS ENZYMATIC ACTIVITY TOWARDS CRYSTALLINE CELLULOSE.  
 CC AT LONG REACTION TIMES. IT IS ALSO ABLE TO DEGRADE CARBOXYMETHYL  
 CC CELLULOSE AND BARLEY B-GLUCAN.  
 CC -1- CATALYTIC ACTIVITY: Hydrolysis of 1,4-beta-D-glucosidic linkages  
 CC in cellulose and cellotetraose, releasing cellobiose from the non-  
 CC reducing ends of the chains.  
 CC -1- SIMILARITY: BELONGS TO CELLULOSE FAMILY B (FAMILY 6 OF GLYCOSYL

```

CC CC HYDROLASES).
CC -1- SIMILARITY: CONTAINS 1 FUNGAL-TYPE CELLULOSE-BINDING DOMAIN (CBD).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; L24519; AA050607.1; -.
CC EMBL; L24520; AA050608.1; -.
CC EMBL; Z34007; CAA83971.1; -.
CC HGSP; P00725; IAZ6.
CC InterPro; IPR000254; CBD_fungal.
CC Pfam; PF00734; CBM_1; 1.
CC Pfam; PF01341; Glyco_hydro_6; 1.
CC PRINTS; PR00733; GLHYDRLASE6.
CC ProDom; PD001821; CBD_fungal; 1.
CC ProDom; PD003733; GH_6; 1.
CC SMART; SM00236; fCBD; 1.
CC PROSITE; PS00562; CBD_FUNGAL; 1.
CC PROSITE; PS00655; GLYCOSYL_HYDROL_F6_1; 1.
CC PROSITE; PS00656; GLYCOSYL_HYDROL_F6_2; 1.
CC Cellulose degradation; Hydrolase; Glycosidase; Signal.
FT SIGNAL 1 20
FT CHAIN 21 438
FT DOMAIN 21 59
FT DOMAIN 60 87
FT DOMAIN 88 438
FT ACT_SITE 215 215
FT ACT_SITE 393 393
FT DISULFID 28 45
FT DISULFID 39 55
FT DISULFID 170 229
FT DISULFID 360 407
FT VARIANT 133 133
FT VARIANT 152 152
FT VARIANT 244 244
FT VARIANT 248 248
FT VARIANT 398 398
SQ SEQUENCE 438 AA; 46209 MW; 002C973544893794 CRC64;
Query Match
Best Local Similarity 10.2%; Score 192.5; DB 1; Length 438;
Matches 61; Conservative 26; Mismatches 84; Indels 33; Gaps 8;
Qy 9 ALLLAASST---YAAECGQYGGCGKWTGPTCTSGFTCVGNENWYSOCIPNDVQ 65
Db 5 AALLALASLVGFVQAOQSPVWGQCGNGWTGPTTCASGSTCV--KQNDYFSQCLPNNQA- 61
Qy 66 GNPXTTITTT--TTKAAITTKAPVTTKA-----TTTTTKAP-----VTITKATTTTTK 113
Db 62 --PSITTTQCTTPPATTTSGCTGPTSGAGNPTGKTVLWSPFVADVAQAADINPSL 119
Qy 114 TTKTKTTTKATTTSSNTGSPISGFGSGNGRTRVW-----DCKKPSCAWDG 162
Db 120 ATKAASVAKITPFVWFDTVAKVPDLGGYGLADARSKNQLVQIVVYDLPDRDCA--ALASNG 177
Qy 163 KASYTKFVLTKADGVSRLGSDVQ 186
Db 178 EFSLANDGLNKNYVDQIAAQIK 201
RESULT 8
SGS3 DROME
ID SGS3 DROME STANDARD; PRT; 307 AA.
AC P02840; QSVTJ2;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)

```

```

DE DE Salivary glue protein Sgs-3 precursor.
GN GN SGS3 OR CG11720.
OS OS Drosophila melanogaster (Fruit fly).
OC OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
OC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
OC Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
OX OX NCBI_TaxID=7221;
RN RN [1]
RP RP SEQUENCE FROM N.A.
RX RX MEDLINE=83294545; PubMed=6411930;
RA RA Garfinkel M.D., Pruitt R.E., Meyerowitz E.M.;
RT RT "DNA sequences, gene regulation and modular protein evolution in the
RL RL Drosophila 68C glue gene cluster.";
RN RN J. Mol. Biol. 168:765-789(1983).
RN RN [2]
RP RP SEQUENCE FROM N.A.
RX RX STRAIN=Berkely;
RN RN MEDLINE=20196006; PubMed=10731132;
RA RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Ananides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktarglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Fostler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jaisi M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M., M.G.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissenbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT RT "The genome sequence of Drosophila melanogaster.";
RL RL Science 287:2185-2195(2000).
RN RN [3]
RP RP SEQUENCE OF 1-28 FROM N.A.
RX RX MEDLINE=88332966; PubMed=3138416;
RA RA Martin C.H., Mayeda C.A., Meyerowitz E.M.;
RT RT "Evolution and expression of the Sgs-3 glue gene of Drosophila.";
RN RN J. Mol. Biol. 201:273-287(1988).
RN RN [4]
RP RP DEVELOPMENTAL STAGE.
RX RX MEDLINE=94038699; PubMed=8223281;
RA RA Huet F., Ruiz C., Richards G.;
RT RT "Puffs and PCR: the in vivo dynamics of early gene expression during
RL RL ecdysone responses in Drosophila.";
CC CC Development 118:613-627(1993).
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- TISSUE SPECIFICITY: SALIVARY GLAND SPECIFIC.
CC -1- DEVELOPMENTAL STAGE: IN THE SALIVARY GLANDS OF MID INSTAR LARVAE
CC LEVELS DRAMATICALLY INCREASE DURING PUFF STAGE 1 AT 98-106 HOURS

```



CC OF DEVELOPMENT. LEVELS REMAIN CONSTANT AND ABUNDANT IN LATE LARVAE  
CC UNTIL PUPF STAGE 10, THEN DECREASE BY STAGE 11.

CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).

CC -----  
CC EMBL; X01918; CAA25994.1; -;  
CC EMBL; AE003544; AAF50056.1; -;  
CC EMBL; X78392; CAA55154.1; -;  
CC PIR; A03329; GSFE3;  
CC FlyBase; FBgn0003373; Sgs3.  
CC Repeat; Signal.  
CC SIGNAL 1 23 POTENTIAL.  
CC FT CHAIN 24 307 SALIVARY GLUE PROTEIN SGS-3.  
CC SQ SEQUENCE 307 AA; 32196 MW; 45803DED16C418BC CRC64;

Query Match 9.9%; Score 187; DB 1; Length 307;  
Best Local Similarity 35.5%; Pred. No. 1.4e-06;  
Matches 66; Conservative 17; Mismatches 59; Indels 44; Gaps 11;

QY 1 MKFSI---IASALLAASSTYAAECSSQYGGQGWMTGTCCTSGFTCVGAENNEWYSQ 57  
DB 1 MKLTATLASILLIGAN--VANCC-----DCG-----CPTTTTCA----- 36  
QY 58 CIPNDVQGNPKTTTTTTTAA-----TTTKAPVTTTKAAT--TTTKAPVTTTKAATTT 110  
DB 37 --PRTPPPCTTTTTTTTCAAPPTQSTTQPCTTSKPTTKQTTLQTPCTTPTTKAT 94  
QY 111 TTK-TTKTTTKAAT---TTSSNTGYSPISGFGNGRTRYWDCCKPSCAWDGKASV 166  
DB .95 TTKPTTKAATTKAATTKPTTKQTTLQTPCTTP-TTKQTTLQTPCTTPTTT---KPTT 150  
QY 167 TKPVL 172  
DB 151 TKPTT 156

RESULT 9  
SGS3 DROSI STANDARD; PRT; 217 AA.

AC F13729;  
DT 01-JAN-1990 (Rel. 13, Created)  
DT 01-AUG-1991 (Rel. 19, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Salivary glue protein Sgs-3 precursor.

GN SGS3.  
OS Drosophila simulans (Fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;  
OC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;  
OC Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.  
OX NCBI\_TaxID=7240;

RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=88332966; PubMed=3138416;  
RA Martin C.H., Mayeda C.A., Meyerowitz E.M.;  
RT "Evolution and expression of the Sgs-3 glue gene of Drosophila.";  
RL J. Mol. Biol. 201:273-287(1988).  
CC -I- DEVELOPMENTAL STAGE: PRODUCED BY THIRD-INSTAR LARVAE.  
DR PIR; S01358.  
DR FlyBase; FBgn0012853; Dsim|Sgs3.  
KW Repeat; Signal.

FT SIGNAL 1 23  
FT CHAIN 24 217 SALIVARY GLUE PROTEIN SGS-3.  
SQ SEQUENCE 217 AA; 22750 MW; D29894E340257881 CRC64;

Query Match 9.6%; Score 181.5; DB 1; Length 217;  
Best Local Similarity 31.7%; Pred. No. 2.5e-06;  
Matches 71; Conservative 20; Mismatches 68; Indels 65; Gaps 13;

QY 4 STIASALLAASSTYAAECSSQYGGQGWMTGTCCTSGFTCVGAENNEWYSQIPNDQ 63  
DB 7 TVLASILLIGFAN--VANCS-----DCG-----CPTKAIT--TCAPPTK-----PTCK 45  
QY 64 VQGNPKTTTTTTTAAATTTKAPVT---TTKATTTTTTKAP-----VTTTKAT 107  
DB 46 STSTTTTTTTTTTTTTTTTTRAPPTKPTCKSTSTTTTTTTRAPPTKPTCKSTSTTTTTTRAP 105  
QY 108 TTTTTKTTTKTTTAAATTTSSNTGYSPISGFGNGRTRYWDCCKPSCAWDGKASV 167  
DB 106 PTTTCKTSTTTTTTHKPTTHSPKT--KP-----TKHTPTKPTKHTPTKPTKPT 153  
QY 168 K---PVLF-----CAKQGVSRIGSDVQSGVGVGOAYMCD 199  
DB 154 KHTTPTTTTTTTPKPGCKSCGP---GGEPCGKC-GKRSALCQD 193

RESULT 10

SGS3 DROVA

ID SGS3 DROVA STANDARD; PRT; 263 AA.

AC F13728;

DT 01-JAN-1990 (Rel. 13, Created)

DT 01-AUG-1991 (Rel. 19, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Salivary glue protein Sgs-3 precursor.

GN SGS3.

OS Drosophila yakuba (Fruit fly).

OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;

OC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;

OC Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

OX NCBI\_TaxID=7245;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=88332966; PubMed=3138416;

RA Martin C.H., Mayeda C.A., Meyerowitz E.M.;

RT "Evolution and expression of the Sgs-3 glue gene of Drosophila.";

RL J. Mol. Biol. 201:273-287(1988).

CC -I- DEVELOPMENTAL STAGE: PRODUCED BY THIRD-INSTAR LARVAE.

DR PIR; S01360; S01360.

DR FlyBase; FBgn0013172; Dyak|Sgs3.

KW Repeat; Signal.

FT SIGNAL 1 23

FT CHAIN 24 263 SALIVARY GLUE PROTEIN SGS-3.

SQ SEQUENCE 263 AA; 28392 MW; C0C5246B482A261C CRC64;

Query Match 9.6%; Score 181.5; DB 1; Length 263;

Best Local Similarity 37.9%; Pred. No. 3e-06;

Matches 58; Conservative 14; Mismatches 50; Indels 31; Gaps 7;

QY 1 MKFSI---IASALLAASSTYAAECSSQYGGQGWMTGTCCTSGFTCVGAENNEWYSQ 57

DB 1 MKLTATLASILLIGAN--VANCC-----DCG-----CPTTTTCA----- 36

QY 58 CIPNDVQGNPKTTTTTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTT 54

DB 55 CAPTRPPPPCTDAPTTTKRTEKSTRTTTRTTTTRPTTTT--TTTTTRPTTR 112

QY 102 -TTTKATTTTTTKT---TKTTTTKAAATTTSS 130

DB 113 STTRHTTTTTTTTTTTRRPTTTTTTTTTTTRPTTTTTT 145

RESULT 11

VGLX HSVB

ID VGLX HSVB STANDARD; PRT; 797 AA.

AC P28968;

DT 01-DEC-1992 (Rel. 24, Created)

DT 01-DEC-1992 (Rel. 24, Last sequence update)

DT 01-DEC-1992 (Rel. 24, Last annotation update)

DE Glycoprotein X precursor.

GN 71.

OS Equine herpesvirus type 1 (strain Ab4p) (EHV-1).



QY 267 GCGVGFGVNGCQSQWNTNTDGMGARYGGISSISBCDKLPTOL 307  
 DB 273 G-----NDWQSAGAFISGSAA---ALSOVTNPDGSGTTNL 304

## RESULT 13

MUC2 HUMAN  
 ID MUC2 HUMAN STANDARD; PRT; 5179 AA.  
 AC Q02817; Q14878;  
 DT 01-JUN-1994 (Rel. 29, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Mucin 2 precursor (intestinal mucin 2).  
 GN MUC2 OR SMUC.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Intestine;  
 RX MEDLINE=941132002; PubMed=8300571;  
 RA Gum J.R. Jr., Hicks J.W., Toribara N.W., Siddiki B., Kim Y.S.;  
 RT "Molecular cloning of human intestinal mucin (MUC2) cDNA.  
 RT Identification of the amino terminus and overall sequence similarity  
 RT to prepro-von Willebrand factor";  
 RL J. Biol. Chem. 269:2440-2446(1994).  
 RN [2]  
 RP SEQUENCE OF 626-1895 AND 4196-5179 FROM N.A.  
 RC TISSUE=Colon;  
 RX MEDLINE=93016075; PubMed=1400449;  
 RA Gum J.R. Jr., Hicks J.W., Toribara N.W., Rothe E.-M., Lagace R.E.,  
 RA Kim Y.S.;  
 RT "The human MUC2 intestinal mucin has cysteine-rich subdomains located  
 RT both upstream and downstream of its central repetitive region.";  
 RL J. Biol. Chem. 267:21375-21383(1992).  
 RN [3]  
 RP SEQUENCE OF 1343-1895 AND 4176-4195 FROM N.A.  
 RX MEDLINE=91358717; PubMed=1885763;  
 RA Toribara N.W., Gum J.R. Jr., Culhane P.J., Lagace R.E., Hicks J.W.,  
 RA Petersen G.N., Kim Y.S.;  
 RT "MUC-2 human small intestinal mucin gene structure. Repeated arrays  
 RT and polymorphism";  
 RL J. Clin. Invest. 88:1005-1013(1991).  
 CC -1- FUNCTION: COATS THE EPITHELIA OF THE INTESTINES, AIRWAYS, AND  
 CC OTHER MUCUS MEMBRANE-CONTAINING ORGANS. THOUGHT TO PROVIDE A  
 CC PROTECTIVE, LUBRICATING BARRIER AGAINST PARTICLES AND INFECTIOUS  
 CC AGENTS AT MUCOSAL SURFACES.  
 CC -1- SUBUNIT: MULTIMERIC.  
 CC -1- TISSUE SPECIFICITY: COLON, SMALL INTESTINE, COLONIC TUMORS,  
 CC BRONCHUS, CERVIX AND GALL BLADDER.  
 CC -1- PTM: ALL CYSTEINE RESIDUES ARE INVOLVED IN INTRACHAIN OR  
 CC INTERCHAIN DISULFIDE BONDS (BY SIMILARITY).  
 CC -1- POLYMORPHISM: THE NUMBER OF REPEATS IS HIGHLY POLYMORPHIC AND  
 CC VARIES AMONG DIFFERENT ALLELES.  
 CC -1- SIMILARITY: THE N-TERMINAL DOMAIN SHOWS SOME SIMILARITY TO THAT  
 CC OF SILKWORM HEMOCYTIN.  
 CC -1- SIMILARITY: CONTAINS 2 VNFC DOMAINS.  
 CC -1- SIMILARITY: CONTAINS 1 C-TERMINAL CYSTEINE KNOT-LIKE (CTCK) DOMAIN.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; L21998; AAB95295.1; -  
 CC EMBL; M74027; AAA59875.1; -  
 CC EMBL; M94131; AAA59163.1; -

DR EMBL; M94132; AAA59164.1; -  
 DR Genew; HGNC:7512; MUC2.  
 DR MIM; 158370; -  
 DR InterPro; IPR000359; Cys\_knot.  
 DR InterPro; IPR000561; EGF-like.  
 DR InterPro; IPR002400; GF\_cysknot.  
 DR InterPro; IPR002919; TIL\_Cysrich.  
 DR InterPro; IPR001007; VWF\_C.  
 DR InterPro; IPR001846; VWF\_D.  
 DR Pfam; PF000093; vwc; 1.  
 DR Pfam; PF00094; vwd; 4.  
 DR Pfam; PF01826; TIL; 1.  
 DR PRINTS; PR00438; GFCYSKNOT.  
 DR SMART; SM00214; VWC; 2.  
 DR SMART; SM00216; VWD; 4.  
 DR PROSITE; PS00022; EGF\_1; UNKNOWN\_1.  
 DR PROSITE; PS01185; CTCK\_1; 1.  
 DR PROSITE; PS01225; CTCK\_2; 1.  
 DR PROSITE; PS01208; VWC; 2.  
 KW Glycoprotein; Repeat; Signal.  
 FT SIGNAL 1 20 POTENTIAL.  
 FT CHAIN 21 5179 MUCIN 2.  
 FT DOMAIN 1401 1747 APPROXIMATE REPEATS.  
 FT REPEAT 1401 1416 1.  
 FT REPEAT 1417 1432 2.  
 FT REPEAT 1433 1448 3.  
 FT REPEAT 1449 1464 4.  
 FT REPEAT 1465 1471 5.  
 FT REPEAT 1472 1478 6.  
 FT REPEAT 1479 1494 7A.  
 FT REPEAT 1495 1517 7B.  
 FT REPEAT 1518 1533 8A.  
 FT REPEAT 1534 1556 8B.  
 FT REPEAT 1557 1572 9A.  
 FT REPEAT 1573 1596 9B.  
 FT REPEAT 1597 1612 10A.  
 FT REPEAT 1613 1635 10B.  
 FT REPEAT 1636 1651 11A.  
 FT REPEAT 1652 1675 11B.  
 FT REPEAT 1676 1683 12.  
 FT REPEAT 1684 1699 13.  
 FT REPEAT 1700 1715 14.  
 FT REPEAT 1716 1731 15.  
 FT REPEAT 1732 1747 16.  
 FT DOMAIN 4815 4886 VNFC 1.  
 FT DOMAIN 4924 4991 CTCK.  
 FT DOMAIN 5075 5122 BY SIMILARITY.  
 FT DISULFID 5075 5122 BY SIMILARITY.  
 FT DISULFID 5089 5136 BY SIMILARITY.  
 FT DISULFID 5098 5152 BY SIMILARITY.  
 FT DISULFID 5102 5154 BY SIMILARITY.  
 FT DISULFID ? 5159 BY SIMILARITY.  
 FT CARBOHYD 163 163 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 423 423 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 670 670 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 770 770 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 894 894 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 1139 1139 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 1154 1154 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 1215 1215 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 1230 1230 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 1246 1246 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 1787 1787 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 1820 1820 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 4339 4339 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 4351 4351 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 4362 4362 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 4373 4373 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 4422 4422 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 4438 4438 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 4502 4502 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 4616 4616 N-LINKED (GLCNAC. .) (POTENTIAL).

FT	CARBOHYD	4627	4627	N-LINKED (GLCNAC. .)	(POTENTIAL).
FT	CARBOHYD	4752	4752	N-LINKED (GLCNAC. .)	(POTENTIAL).
FT	CARBOHYD	4787	4787	N-LINKED (GLCNAC. .)	(POTENTIAL).
FT	CARBOHYD	4881	4881	N-LINKED (GLCNAC. .)	(POTENTIAL).
FT	CARBOHYD	4898	4898	N-LINKED (GLCNAC. .)	(POTENTIAL).
FT	CARBOHYD	4955	4955	N-LINKED (GLCNAC. .)	(POTENTIAL).
FT	CARBOHYD	4970	4970	N-LINKED (GLCNAC. .)	(POTENTIAL).
FT	CARBOHYD	5019	5019	N-LINKED (GLCNAC. .)	(POTENTIAL).
FT	CARBOHYD	5038	5038	N-LINKED (GLCNAC. .)	(POTENTIAL).
FT	CARBOHYD	5069	5069	N-LINKED (GLCNAC. .)	(POTENTIAL).
FT	CARBOHYD	1351	1351	H -> L (IN REF. 3).	
FT	CONFLICT	1412	1412	T -> S (IN REF. 3).	
FT	CONFLICT	1449	1449	L -> P (IN REF. 3).	
FT	CONFLICT	1504	1504	M -> T (IN REF. 3).	
FT	CONFLICT	4192	4192	G -> S (IN REF. 2).	
FT	CONFLICT	4192	4192	G -> S (IN REF. 2).	
FT	SEQUENCE	5179 AA;	540295 MW;	85CD7571PB9A5663 CRC64;	
QY	Query Match	8.7%;	Score 164;	DB 1:	Length 5179;
QY	Best Local Similarity	31.7%;	Pred. No. 0.00086;		
QY	Matches	57;	Conservative	12;	Mismatches 65; Indels 46; Gaps 8;
Db	QY	67	NPXTTITTTTAAATTKAPVTTTAKATTTTTTTKAPVTITTKAT--TTTTTKTTTKTITTKAA	124	
Db	1690	SPFITWTTPSPPTTSSPIITTTTSSITTSPPTTTPSPPPTMTTTPSTPTTTPSPPPTMTTLPP	1749		
QY	125	TTTTSSNNT-----GYSPISGFGSGRTTRYWDCCKPSCAW-----DGKASVTK	168		
Db	1750	TTTTSSPLTTTLPSPSIPTTFSPFS-----TTTTPTTCVCLCWNWTGMDSGKNPFHK	1801		
QY	169	PVLTCARDGVSRSLGSDDVQSCGCGQAQYMCNDNOPVVNDDL-AYGFAAAALSGSAGASAFC	227		
Db	1802	P-----GGDTLIGDV-----CGPG-----WAANISCRATMYDPVIGQLGTVCV	1842		
RESULT 14					
DNA4 YEAST					
ID	DNA4 YEAST	STANDARD;	PRT;	1161 AA.	
AC	P47179;				
DT	01-FEB-1996 (Rel. 33, Created)				
DT	01-FEB-1996 (Rel. 33, Last sequence update)				
DT	15-JUN-2002 (Rel. 41, Last annotation update)				
DE	Cell wall protein DNA4 precursor.				
GN	DNA4 OR YJR151C OR J2223.				
OS	Saccharomyces cerevisiae (Baker's yeast).				
OC	Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;				
CC	Saccharomycetales; Saccharomycetaceae; Saccharomycetes.				
ON	NCBI_TaxID=4932;				
OX	[1]				
RP	SEQUENCE FROM N.A.				
RA	Scarcez T.;				
RL	Submitted (SEP-1995) to the EMBL/GenBank/DDBJ databases.				
RN	[2]				
RP	REGULATION.				
RX	MEDLINE=211131168; PubMed=11160904;				
RA	Cohen B.D., Serfil O., Abramova N.E., Davies K.J., Lowry C.V.;				
RT	"Induction and repression of DAN1 and the family of anaerobic				
RT	mammoprotein genes in Saccharomyces cerevisiae occurs through a				
RT	complex array of regulatory sites."				
RT	Nucleic Acids Res. 29:799-808(2001).				
CC	-!- FUNCTION: COMPONENT OF THE CELL WALL (By similarity).				
CC	-!- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor				
CC	(Potential).				
CC	-!- PTM: EXTENSIVELY O-GLYCOSYLATED (POTENTIAL).				
CC	-!- SIMILARITY: BELONGS TO THE SRP1 / TIPI FAMILY.				
CC	-----				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration				
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation				
CC	the European Bioinformatics Institute. There are no restrictions on its				
CC	use by non-profit institutions as long as its content is in no way				
CC	modified and this statement is not removed. Usage by and for commercia				
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce				
CC	or send an email to license@isb-sib.ch).				
CC	-----				

```

DR EMBL; Z49651; CAA89604.1; -.
DR SGD; S0003912; DAN4.
DR InterPro; IPR000992; SRP1_TIP1.
DR Pfam; PF00660; SRP1_TIP1; 1.
DR PROSITE; PS00724; SRP1_TIP1; 1.
KW Cell wall; Glycoprotein; Membrane; GPI-anchor; Signal.
FT SIGNAL 1 24 POTENTIAL.
FT CHAIN 25 1146 CELL WALL PROTEIN DAN4.
FT PROPSP 1147 1161 REMOVED IN MATURE FORM (POTENTIAL).
FT LPIP 1146 1146 GPI-ANCHOR (POTENTIAL).
SQ SEQUENCE 1161 AA; 118358 MW; 7954C15D89FOCA58 CRC64;
Query Match 8.6%; Score 163; DB 1; Length 1161;
Best Local Similarity 36.5%; Pred. No. 0.00025;
Matches 38; Conservative 21; Mismatches 35; Indels 10; Gaps 3;
QY 54 WY-----SQCIPNDQV-QGNPKTTTTTTTAAATTKAPVTTKATTITTKAPVT 104
DB |||||:||||:||||:||||:||||:||||:||||:||||:||||:||||: 157
99 WTSLKPAISSALSKDGIYTAIPTSTSTTTKSSTST-TPTTITSTTSTTPTTST 157
QY 105 KAITTTTKTKTTTAAATTTSSNTGSPISGGFSGNGRTT 148
DB :|||:||:||||:||||:||||:||||:||||:||||:||||:||||: 157
158 TSTPTTSTTSTTPTTSTTPTTSTTPTTSTTPTTSTTPTTSTTPTTSTTPTTS 201
RESULT 15
MUAL_XENLA STANDARD; PRT; 400 AA.
ID MUAL_XENLA AC
AC P10667;
DT 01-JUL-1989 (Rel. 11, Last sequence update)
DT 01-JUL-1989 (Rel. 11, Last sequence update)
DT 01-OCT-1994 (Rel. 30, Last annotation update)
DE Intersegmentary mucin A.1 precursor (FTM-A.1) (Preprospasmodysyn).
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae;
OC Xenopodinae; Xenopus.
OX NCBI_TaxID=83355;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=88227968; PubMed=3372504;
RA Hofmann W.;
RT "A new repetitive protein from Xenopus laevis skin highly homologous
RT to pancreatic zymolytic polypeptide."
RL J. Biol. Chem. 263:7686-7690(1988).
RN [2]
RP CHARACTERIZATION.
RX MEDLINE=90316191; PubMed=2196180;
RA Hauser F., Gertzen E.M., Hoffmann W.;
RT "Expression of spasmodysyn (FTM-A.1); an integumentary mucin from
RT Xenopus laevis."
RL Exp. Cell Res. 189:157-162(1990).
CC -! FUNCTION: COULD BE INVOLVED IN DEFENSE AGAINST MICROBIAL
CC INFECTONS. PROTECTS THE EPITHELIA FROM EXTERNAL ENVIRONMENT.
CC -! SUBCELLULAR LOCATION: Secreted
CC -! TISSUE SPECIFICITY: EXPRESSED AND STORED EXCLUSIVELY IN MATURE
CC MUCOUS GLANDS OF THE SKIN.
CC -! PTM: EXTENSIVELY O-GLYCOSYLATED. CONSIST OF ABOUT 70% CARBOHYDRATE
CC AND 30% PROTEIN.
CC -! SIMILARITY: CONTAINS A P-TYPE (TREFOIL) DOMAINS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC at the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to licens@isb-sib.ch).
CC EMBL; M19971; AAA49960.1; -.
DR PIR; A28172; A28172.
DR HSPSP; P01359; 2FSP.
DR InterPro; IPR000519; P_trefoil.

```



**THIS PAGE BLANK (USPTO)**